DOI：10.13998/j.cnki.issn1002-1248.23-0699

科学数据价值增值影响因素系统动力学仿真研究

孙俐丽[1]，王伟杰[2]，盛杰霏[3]
（1. 南京工业大学 信息管理与技术研究所，南京 210009；2. 南京传媒学院，南京 211172；
3. 南京工业大学 经济与管理学院，南京 211816）

摘　要：[目的 / 意义]探究科学数据价值增值影响因素，揭示科学数据价值增值内在发展规律。[方法 / 过程]基于专家访谈资料和文献中科学数据价值增值相关内容，采用扎根理论方法，得到科学数据价值增值影响因素的理论模型。在此基础上，建立科学数据价值增值系统动力学模型，对科学数据价值增值的过程进行动态模拟仿真与分析，揭示各影响因素对科学数据价值增值的作用关系。[结果 / 结论]科学数据价值增值过程中，原始数据质量是前提条件，数据存缴水平影响显著，数据组织与整合是价值增值形成的关键，科学数据共享与开发是价值增值实现的关键。

关键词：科学数据；价值增值；影响因素；动力学仿真；数据要素；FAIR
中图分类号：G203　　文献标识码：A　　文章编号：1002-1248（2023）09-0028-15

引用本文：孙俐丽，王伟杰，盛杰霏. 科学数据价值增值影响因素系统动力学仿真研究[J]. 农业图书情报学报，2023，35(9)：28-42.

1 引　言

在数据密集型科研范式背景下，科学数据成为驱动科技进步与知识创新的重要基石，逐渐成为各国科技竞争的新战略高地，科学数据价值受到前所未有的关注与重视。《中共中央国务院关于构建更加完善的要素市场化配置体制机制的意见》进一步明确提出，“加快培育数据要素市场” “提升数据资源价值”，为挖掘科学数据价值指明方向。

开放科学运动的发展，推动了科学数据开放共享实践趁势而兴。为了提高科学数据开放水平，国际上许多国家和地区围绕科学数据开放共享制定了若干政策、法规和指南。如美国科学与技术政策办公室发布的《促进科学研究成果获取备忘录》[1]、欧盟《开放数据和公共信息再利用指令》[2]、中国的《科学数据管理办法》等，这些政策要求公共资助产生的科学数据实行公开为常态，不公开为例外的原则面向社会提供开放共享。为了推进科学数据的规范管理进而提高科学数据应用价值，国际社会于 2014 年在荷兰莱顿举办的学术会议上提出了科学数据 FAIR（可发现、可访问、可互操作和可重用）原则草案[3]。FAIR 原则自提出后

收稿日期：2023-08-07
基金项目：江苏省社会科学基金项目“基于共生理论的江苏省科学大数据开放共享机制与模式研究”（19TQC001）
作者简介：孙俐丽（1984- ），博士，副教授，硕士生导师，南京工业大学信息管理与技术研究所，研究方向为科学数据管理、知识产权信息服务。王伟杰（1997- ），硕士，南京传媒学院，助教，研究方向为科学数据管理。盛杰霏（1997- ），女，硕士，南京工业大学经济与管理学院，研究方向为科学数据管理

得到国际社会的广泛认可，许多国家和组织将其运用于科学数据开放共享实践，为促进科学数据质量提升提供有力支撑。在基础设施方面，各国不断升级科学数据共享平台，已形成了综合型、领域型、机构型等多种类型科学数据平台，如美国的“开放科学中心”（COS）、澳大利亚的“研究数据共享基础设施”（ARDC）项目、欧洲的“欧洲开放科学云”（EOSC）项目，中国各学科领域学科数据共享基础设施也逐渐完善，目前已形成 20 个国家科学数据中心[4]。这些科学数据共享平台的搭建，为科学数据价值的实现提供了有力抓手。在科学数据管理服务方面，国外许多高校已经建立起集中的科研数据管理（Research Data Management，RDM）服务来为研究人员提供支持，如英国剑桥大学提出的数据拥护者计划，通过在部门或学院设立本地的数据拥护者来推广良好的数据管理实践[5]。总体而言，目前国内外科学数据开放共享实践主要围绕政策制定、平台建设、数据管理服务等方面展开，这些实践扩大了科学数据开放规模、提升了科学数据质量，有力地推动了科学数据价值实现。

作为开放科学关键要素，科学数据开放共享已形成国际共识，科学数据资源建设、管理与开发工作初见成效。但总体而言，当前科学数据管理普遍存在“重数量轻质量、重开放轻增值”现象[6]，在科学数据质量、科学数据重用等方面面临越来越严峻的挑战，严重影响科学数据要素作用的发挥和科学数据价值实现。科学数据价值问题是数据管理的本源问题，是支配科学数据开放共享的主线[7]。因此，在科学数据开放共享实践发展的当下，围绕科学数据价值增值这一核心问题展开探讨恰逢其时。

2 相关研究成果

现有数据价值的研究主要关于数据价值形成或实现。从经济学的角度，李海建等认为数据质量程度、加工程度、使用程度、连接程度、应用场景、开放程度等影响数据价值形成[8]。从技术角度，JOAO 认为利用云计算等开放技术处理数据，能够提升开放数据价值[9]；从数据组织与加工角度，PETR 等认为大数据的可发现程度及元数据质量影响大数据价值[10]；李旭晖等提出大数据的价值源于其蕴含的各种知识关联，大数据价值的核心在于对知识关联进行刻画、揭示和利用[11]。关于数据要素价值的实现，马费成等认为数据采集是数据要素价值实现源头，数据组织是数据价值开放的重要环节，数据流通是关键，数据利用是价值实现的“最后一公里”[12]；夏义堃等着重考虑数据要素主体、客体与环境 3 个维度探讨了数据要素价值影响因素，并在此基础上提出了数据要素价值提升路径[13]。

关于科学数据价值研究，学者们主要围绕价值标准、价值实现策略等展开理论探讨。邓君等提出了科学数据价值鉴定的原则与标准[14]。顾立平等从图书馆业务实践角度，提出科学数据价值提升的理论框架[15]。孙建军等提出科学大数据价值链应以科学大数据资源为基、以科学大数据管理为力、以科学大数据挖掘为核、以科学大数据服务为标[16]。冯媛基于价值共创理论视角，分析了科学数据开放共享的价值流和多元主体互动关系，提出价值共创模型[17]。任颖等构建了科学数据价值共创系统，认为科学数据价值共创主体和其系统互动程度，以及政治、经济、文化和技术等环境因素共同影响科学数据价值共创[18]。

总体而言，科学数据开发利用与价值实现一直是学界关注热点。现有相关研究大多从科学数据价值主体、客体、环境等维度探讨科学数据价值实现，且以静态的、理论分析为主。科学数据价值增值是价值实现的进一步延伸，涉及科学数据全生命周期过程，需要动态地、全局性地剖析科学数据价值增值规律。因此，本研究运用扎根理论识别科学数据价值增值影响因素，并运用动力学方法进行仿真分析，通过模拟影响因素的作用过程及效应，揭示科学数据价值增值内在规律。

3 科学数据价值增值内涵及路径分析

3.1 科学数据价值增值概念内涵

现有文献中，顾立平等认为数据价值提升是通过

DOI：10.13998/j.cnki.issn1002-1248.23-0699

著录、标引、建立知识关联等规范的组织和管理使科学数据从无序转为有序，使科学数据中隐藏的信息被挖掘出来，从而创造更大的科研价值[15]。任福兵等提出政府开放数据价值增值是对数据资源的进一步整合与开发，形成个性化的增值产品或服务的活动[19]。显然，前者的价值提升侧重数据资源的组织管理过程，后者的价值增值强调数据资源的开发利用。本研究认为在当前科学数据开放共享背景下，科学数据价值增值应着眼于科学数据的全生命周期，即在科学数据生成、存储、组织、开放共享、开发利用等全生命周期过程中，相关利益主体通过管理手段以及数据加工技术，提升科学数据价值效应、激发科学数据的潜在价值或产生增值数据产品或服务。

3.2 科学数据价值增值路径分析

借鉴信息价值链理论[20]，科学数据价值增值不仅仅是各类数据管理活动的简单叠加，而是相关利益主体紧密协同，促进各参与者之间资源、信息、能力充分流通及创新整合，进而实现科学数据价值增值。开放科学数据背景下，伴随科学数据流由“原始科学数据 - 集中存储的科学数据 - 公开的科学数据 - 增值科学数据与服务”的变化，科学数据价值形态实现了由低价值到高价值的价值增值过程（图 1）。

（1）科学数据生成阶段。科研人员或科研团队等数据生产者在科研过程中产生原始科学数据。原始科学数据质量是后续管理与开发利用的基础，同时，这一阶段科学数据生产者掌控数据，科学数据贮存在数据生产者个人设备中，科学数据的价值效能有限。

（2）科学数据集中存储阶段。原始科学数据或机构内部聚合的科学数据进一步汇交至科学数据中心或科学数据出版发行平台，由科学数据中心等对汇交的科学数据进行集中存储。科学数据从私域走向第三方，科学数据的安全性、可靠性等得到一定程度提升。

（3）科学数据组织整合阶段。原始科学数据或机构内部聚合的科学数据进一步汇交至科学数据中心或科学数据出版发行平台，由专业人员对科学数据进行质量审核，对通过审核的科学数据进行规范化处理，赋予高质量元数据，再将经过深度刻画和揭示的科学数据建立知识关联，然后向公众开放。这一阶段使得科学数据质量、可信度、交互性等得到提升，科学数据价值进一步提高。

（4）科学数据开发阶段。科学数据开放共享后，更多的社会机构和组织可以参与对科学数据的创新整合、开发利用。数据研发机构根据用户需求，采取智能化手段对科学数据进行二次开发，形成增值数据产品。数据运营机构针对市场需求，打造个性化科学数据服务，形成科学衍生价值。

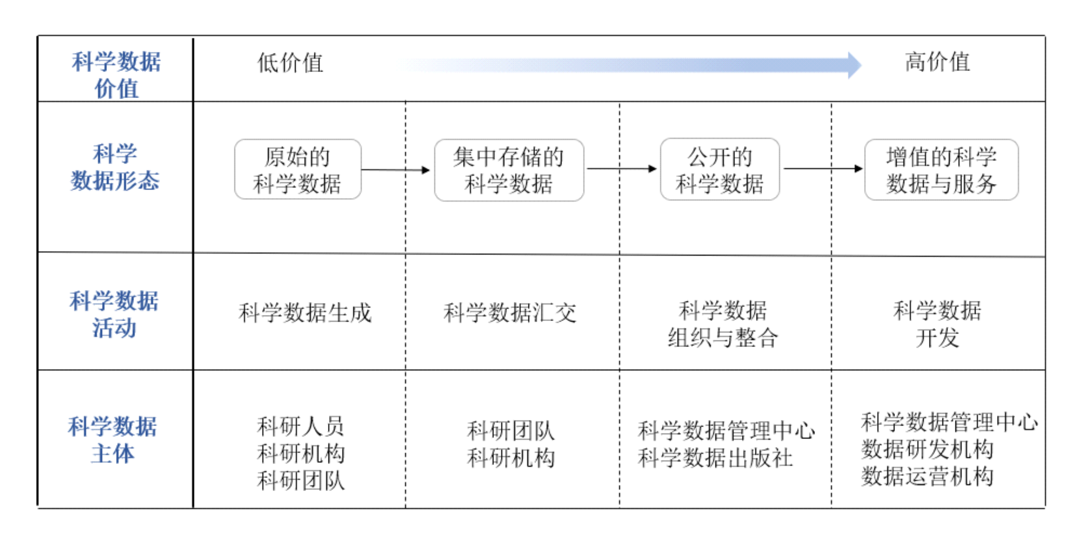


图 1 科学数据价值增值路径

Fig.1 The path of adding value of scientific data

4 科学数据价值增值影响因素识别

4.1 研究方法

已有研究成果中关于科学数据价值增值影响因素直接研究成果较少，缺少可以直接采用或借鉴的成熟理论。本研究将采用扎根理论识别科学数据价值增值的影响因素。扎根理论是一种较为严谨的质性研究方法，适用于从复杂的质性材料中构建理论模型[21]。与量化研究呈相反方向，扎根理论研究是直接从数据着手，逐级抽象和归纳形成理论，是自下而上的研究过程。

4.2 数据获取

扎根理论强调样本资料的全面性和代表性，重视样本资料来源的广度与深度。因此，本研究综合采用专家访谈、文献收集方式获取样本资料。通过对专家经验和已有文献中相关信息的整理、归纳、抽象，最终形成科学数据价值增值影响因素的理论模型。

4.2.1 访谈数据获取

为了避免访谈中可能出现的“霍桑效应”，本研究在访谈过程遵循 MYERS 等提出的规范性访谈框架[22]。为了确保访谈提纲没有歧义，更容易被访谈对象理解，本研究进行了预访谈，并根据预访谈中出现的问题，对访谈提纲的相关表述进行优化。最终访谈提纲主要包含以下问题：①主要通过哪些途径生成 / 采集科研数据？此过程有哪些困难？②是否愿意通过提交至数据中心或数据期刊等方式开放共享科学数据？③科学数据加工组织过程，有哪些关键环节，哪些因素会影响数据价值增值？④科学数据开放利用过程，哪些因素可能影响科学数据价值增值？在访谈开始前，访谈人员需要向访谈对象解释访谈目的和主题，作出保密承诺，并就访谈录音征得对方同意。在访谈过程中，根据访谈对象数据角色不同及访谈具体情境，对具体问题的表述适当调整。

访谈工作主要集中在 2022 年 10—11 月期间，一共选取了具有科学数据生成和利用经历的科研人员 13 名，以及某省科学数据资源统筹中心相关工作人员 5 名参与专家访谈。经访谈对象同意将访谈录音转录，共形成访谈文本 3 万余字。随机抽取其中 13 份访谈文本进行编码分析，其余 5 份留作饱和度检验。

4.2.2 文献数据获取

2022 年 8 月至 2022 年 12 月期间，从 CNKI 及 Web of Science 核心集数据中检索相关文献，最终选取 42 篇与本研究相关的文献，随机选取其中 35 篇进行编码，其余文献留作饱和度检验。

4.3 影响因素识别

（1）开放性编码。首先，对所获得的样本资料反复研读，熟悉将要编码的资料内容，并将原始样本资料逐个分析形成条目数据，完成初始编码。接着，编码小组结合研究主题，通过反复讨论、比较条目所描述的内容与科学数据价值增值的相关性以及条目之间的可区分性等原则，对初始条目进行合并和精简，共形成 76 个条目。将编码人员分为 A、B 编码组，在不准互相沟通的前提下，分别根据条目所体现的内容进行归类和命名。再由 A、B 组编码人员分别阐述分类和命名的理由，经过反复研讨，选择其中获得课题组一致通过的主题概念。最终，通过对获取的初步概念进行分析，将含义相同、相似或相互联系的概念合并于相同范畴之中，形成 20 个范畴。由于篇幅所限，截图部分开放编码过程如表 1 所示。

（2）主轴编码。通过剖析初始范畴的内容，将相互独立的范畴联系起来，发展形成更深层的主范畴。本研究基于对科学数据价值增值的路径分析，结合开放编码结果特点，最终将 20 个范畴归纳为 7 个主范畴。

（3）选择性编码。对主轴式编码所获得的主范畴进行系统分析，选取具有统领性的核心范畴。同时，深入挖掘核心范畴和其他范畴之间的联系，基于此，构建能够反映研究现象本质的理论框架。根据以上思想，本文对形成的 6 个主范畴进行对比剖析，结合“科学数据价值增值”主题，最终确定科学数据价值增值影响因素表。

（4）饱和度检验。理论饱和度是一个主观概念，到目前为止，尚未有度量理论饱和度的客观指标，需

DOI：10.13998/j.cnki.issn1002-1248.23-0699

表 1　部分开放编码过程

Table 1　Some of the open coding processes

部分编码内容	范畴	来源
如果没有强制要求，我不愿意去汇交数据，像我们进行科研主要为了发论文，获得一些成果，如果提交了数据，别人利用我的数据可以抢先发表成果（感知风险）		F2
我本人是愿意去共享或者说去汇交我的数据资源的，但是有时候汇交数据的流程比较繁琐，需要单位层层审核，还需要提交许多汇交材料，就导致我不想去做这件事了（感知努力）	数据汇交意愿	F3
据我了解，目前有的科学数据中心在数据汇交后还会将数据出版或者发布，发布数据就像论文一样，一定程度上也提升了研究人员的名声（感知利益）		F4
就我而言，我觉得汇交数据这件事让我拥有一种成就感，实现了科研人员服务社会的一种责任（感知成就）		F5

要研究者根据研究经验的感性认识进行把握[23]。本研究采用预留的 5 份访谈内容及 7 篇理论文献进行概念比对，结果没有新概念或理论内涵出现，即预留的文本内容基本被此前的 20 个概念所覆盖。根据此种情况本研究认为当前访谈结果及在此基础上形成的理论已经饱和。

（5）信度检验。为了检验上述归类是否恰当，本研究借鉴已有研究的“反向归类法”[24]，请未参与本研究的 3 名研究生在理解上述范畴、主范畴的含义后，将 76 个条目分别归到对应的类别中。反向归类的结果显示，3 位评判者都将该条目归类预想类别中的共有 69 条，有两位评判者将该条目作一致归类的有 7 条，完全不一致的条目为 0。经过深入讨论发现导致未能一致归类的原因主要是对“技术因素”主范畴及其对应的“数智技术应用能力”范畴的区分度存疑，一名评判者认为该项类别的内容已体现在“数据处理能力”“数据开发能力”等范畴中，与这两项类目区分度不足。因此，删除该主范畴及对应范畴，最终得到 19 个范畴，6 个主范畴。此外，随机选取 3 位前述访谈对象对其所描述的科学数据价值增值影响因素进行归属选择，均能正确对应。通过上述反向回归法和反馈法的验证结果表明，以上扎根理论结果具有一定可信度。最终形成科学数据价值增值影响因素如表 2 所示。

5　系统动力学仿真

系统动力学是研究社会系统动态行为的计算机仿真方法，将社会系统模型化，再输入计算机模拟仿真，进行战略和策略的实验[25]。该方法可以处理复杂系统的建模问题，还能量化调节系统各要素之间的关系[26]。科学数据从形成、存储、组织、共享，到开发利用可以视为一个有机联系的系统，是具有系统性、非线性和复杂性等显著特征的动态反馈系统。因此，运用系统动力学构建系统仿真模型，抓住科学数据价值增值的关键环节，剖析其影响因素之间的关键联系和反馈回路。

5.1　系统边界确定

明确系统边界是系统动力学建模的关键。本研究以科学数据价值增值过程为研究对象，围绕科学数据价值增值过程的相关影响因素进行研究。所以，本研究将科学数据价值增值影响因素划入系统边界，包括原始数据质量因素、数据存缴因素、数据组织与整合因素、数据共享和开发因素、数据政策与监管因素、资金保障因素等 6 个子系统，每个子系统所涉及的因素之间存在复杂的动态因果关系。其中数据存缴因素子系统中，数据汇交意愿对其有较大影响，为了进一步探析数据汇交意愿的作用规律，所以将与其相关的感知风险、感知努力、感知利益、感知成就纳入该子系统。

5.2　因果反馈关系分析

本研究通过因果反馈回路图描绘出科学数据价值增值与影响因素间关系形成的动态过程（图 2）。经过 Vensim PLE 运行，得到科学数据价值增值过程的反馈回路，详见表 3。

表 2　科学数据价值增值影响因素

Table 2 Influencing factors of scientific data value appreciation

核心范畴	主范畴	范畴	范畴内涵
科学数据收集因素	原始数据质量	科研人员数据素养	科研人员在数据收集、处理、使用等过程中体现出的能力和素养
		数据软硬件设施	科学数据采集、处理、存储相关的软硬件条件
		数据服务质量	科学数据管理与服务单位提供科学数据相关服务的质量
	数据存缴	数据汇交意愿	科研人员等科学数据生产者将科学数据存缴至数据中心等平台，提供数据共享的愿意程度
		数据审核机制	科学数据中心等数据存储机构对汇交的科学数据按照一定标准规范对数据进行质量审核的机制
		数据存储与维护机制	科学数据中心等数据存储机构对科学数据进行存储、备份、恢复和管理的相关规范和流程
科学数据组织与开发因素	数据组织与整合	元数据质量	科学数据的元数据质量水平
		数据处理能力	科学数据中心等科学数据管理机构对科学数据进行分类、描述、转化等处理活动的能力
		数据整合能力	科学数据中心等科学数据管理机构对不同领域、多维度科学数据建立关联的能力
	数据共享与开发	数据开放规模	科学数据中心等将科学数据对公众开放的范围、对使用者的约束条件等
		数据开放风险	科学数据开放过程中可能产生的负面影响
		数据共享平台质量	科学数据共享平台功能完善程度、易用性等方面的质量表现
		数据共享服务能力	科学数据中心等数据服务机构提供科学数据利用过程相关服务的能力水平
		数据开发能力	科学数据中心等数据服务机构对科学数据进行二次开发形成新的数据产品或服务的能力
保障因素	数据政策与监管	数据政策完善程度	有关科学数据的宏观、微观等多层面政策的完备程度
		数据评价与激励机制	对科学数据质量、科学数据管理服务等方面的评价与激励机制
		数据监管机制	对科学数据进行管理和监督的制度、程序和措施
	资金保障	财政支持力度	政府对科学数据管理活动的资金支持力度
		数据价值增值收益	科学数据中心等通过科学数据价值增值所获得的经济收益

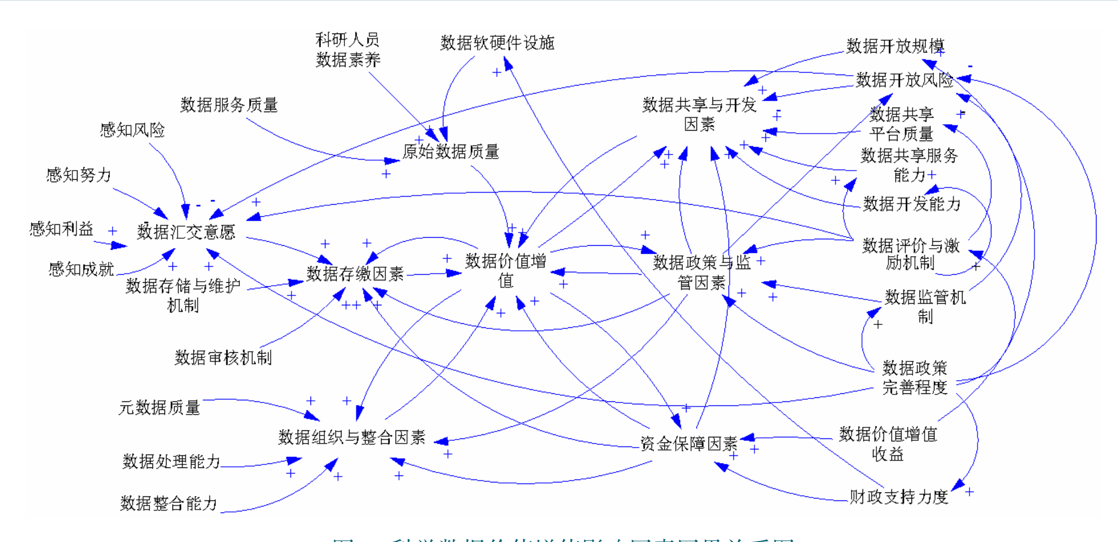


图 2　科学数据价值增值影响因素因果关系图

Fig. 2 Causal diagram of the influencing factors of scientific data value appreciation

5.3 仿真模型构建

5.3.1　系统流图

基于科学数据价值增值影响因素的因果关系，本研究通过 Vensim PLE 工具绘制了科学数据价值增值系统流图，如图 3 所示。该图反映了这 4 类变量之间的相互联系、相互作用方式，揭示了科学数据价值增值过程中各要素之间的作用关系。

DOI：10.13998/j.cnki.issn1002-1248.23-0699

表 3　科学数据价值增值反馈回路

Table 3 Feedback loop of value-added scientific data

回路序号	反馈类型	回路详情
回路 1	正反馈	数据存缴→数据价值增值→资金保障→数据存缴
回路 2	正反馈	数据存缴→数据价值增值→数据政策与监管→数据存缴
回路 3	正反馈	数据共享与开发→数据价值增值→资金保障→数据共享与开发
回路 4	正反馈	数据组织与整合→数据价值增值→资金保障→数据组织与整合
回路 5	正反馈	数据组织与整合→数据价值增值→数据政策与监管→数据组织与整合
回路 6	正反馈	数据共享与开发→数据价值增值→数据政策与监管→数据共享与开发
回路 7	正反馈	资金保障→数据组织与整合→数据价值增值→资金保障
…	…	……

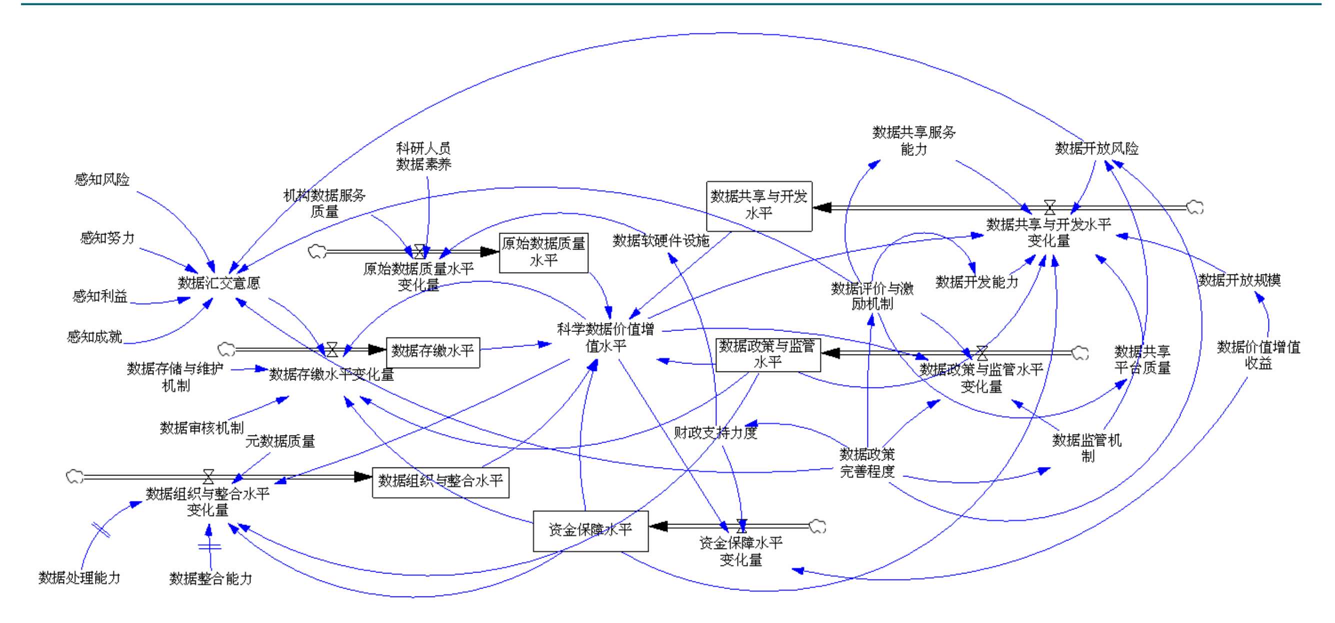


图 3　科学数据价值增值影响因素系统流图

Fig.3 Flow diagram of the value-added system of scientific data

根据图 3 可知，科学数据价值增值系统流图总共涉及 36 个变量，包括 6 个水平变量，6 个速率变量，11 个辅助变量，13 个常量，具体变量与相应的含义如表 4 所示。

5.3.2　基本假设

为了便于数据仿真与模拟，需对科学数据价值增

表 4　变量名称及其类型

Table 4 Variable names and types

变量类型	变量名称
状态变量	原始数据质量水平、数据存缴水平、数据组织与整合水平、数据共享与开发水平、数据政策与监管水平、资金保障水平
速率变量	原始数据质量水平变化量、数据存缴水平变化量、数据组织与整合水平变化量、数据共享与开发水平变化量、数据政策与监管水平变化量、资金保障水平变化量
辅助变量	科学数据价值增值水平、数据软硬件设施、数据汇交意愿、数据开放风险、数据共享服务能力、数据开放规模、数据开发能力、数据共享平台质量、数据监管机制、数据评价与激励机制、财政支持力度
常量	感知风险、感知努力、感知利益、感知成就、数据存储与维护机制、数据审核机制、科研人员数据素养、机构数据服务质量、元数据质量、数据处理能力、数据整合能力、数据政策完善程度、数据价值增值收益

值影响因素系统动力学模型提出基本假设，并对部分参数设置说明。

（1）假设科学数据价值增值影响因素系统是一个各子系统构成要素频繁互动反馈的系统，不考虑系统边界以外的因素对系统的影响。

（2）假设原始数据质量、数据存缴、数据组织与整合等因素会随着实践的积累而发生改变。

（3）假设科学数据价值增值程度可以通过科学数据价值增值水平进行量化。

（4）构建的系统动力学模型具有普适性，不考虑因系统外部因重大自然灾害、重大政策变革等非正常情况或突发事件所引起的系统崩溃。

5.3.3　仿真方程设定与参数设置

科学数据价值增值系统模型运行的前提是对相关常量以及变量的初始值进行设定。本文借鉴高晓宁[27]、袁红[28]等学者的方法，通过问卷调查获得方程所涉及的相关指标权重以及部分变量初始值。问卷采用李克特十级量表，1 到 10 分别表现为从非常不同意到非常同意，通过对涉及的指标变量进行打分，计算各题的平均分作为相应指标变量的影响权重。为了保证研究结论的一致性，调查问卷发放对象是前文研究选取的 18 名访谈人员。问卷通过 QQ、微信、邮件形式进行发送，回收周期为 7 天，回收率 100%。通过对各问卷的分值进行整理，利用层次分析法分析软件 YAAHP 对打分结果进行计算获取各个因素的权重以及整体一致性检验结果，对 18 名专家的打分结果采用算术平均的方法来获取权重。经计算，一致性检验指标 CR＜0，满足研究需求。

最终得到各子系统中影响因素权重：原始数据质量子系统（机构数据服务质量，0.2；科研人员数据素养，0.45；数据软硬件设施，0.35）；数据存缴子系统（数据汇交意愿，0.3；数据存储与维护机制，0.2；数据审核机制，0.25；科学数据价值增值水平，0.05；资金保障水平，0.1；数据政策与监管水平，0.1）；数据组织与整合子系统（元数据质量，0.3；数据处理能力，0.25；数据整合能力，0.2；科学数据价值增值水平，0.05；资金保障水平，0.1；数据政策与监管水平，0.1）；数据共享与开发子系统（数据开发能力，0.14；数据开放规模，0.2；数据共享服务能力，0.13；数据共享平台质量，0.16；数据开放风险，0.12；科学数据价值增值水平，0.05；资金保障水平，0.1；数据政策与监管水平，0.1）；数据政策与监管子系统（数据政策完善程度，0.45；数据评价与激励机制，0.3；数据监管机制，0.2；科学数据价值增值水平，0.05）；资金保障子系统（财政支持力度，0.5；数据价值增值收益，0.45；科学数据价值增值水平，0.05）。

此外，感知风险、感知努力、感知利益、感知成就、数据存储与维护机制、数据审核机制、科研人员数据素养、机构数据服务质量、元数据质量、数据处理能力、数据整合能力、数据政策完善程度、数据价值增值收益等常量这部分数值采用 RANDOM UNIFORM 随机分布函数，并设置数值区间为[0.1]。

模型中涉及的主要变量方程如下：

（1）科学数据价值增值水平 = 原始数据质量水平 × 0.25 + 数据存缴水平 ×0.2 + 数据组织与整合水平 ×0.16 + 数据共享与开发水平 ×0.14 + 数据政策与监管水平 × 0.13 + 资金保障水平 × 0.12

（2）原始数据质量水平 =INTEG（原始数据质量水平变化量，0）

（3）原始数据质量水平变化量 =SMOOTHI（机构数据服务质量 × 0.2 + 科研人员数据素养 × 0.45 + 数据软硬件设施 × 0.35，1，0）

（4）数据存缴水平 =INTEG（数据存缴水平变化量，0）

（5）数据存缴水平变化量 =SMOOTHI（数据汇交意愿 × 0.3+ 数据存储与维护机制 × 0.2 + 数据审核机制 × 0.25+ 科学数据价值增值水平 ×0.05 + 资金保障水平 × 0.1+ 数据政策与监管水平 × 0.1，1，0）

5.4　仿真分析

5.4.1　模型有效性检验

为了保证所构建模型的有效性，确保系统模型动态仿真的顺利开展，本研究对科学数据价值增值的系统动力学模型进行有效性检验，选取初始状态下的科

DOI：10.13998/j.cnki.issn1002-1248.23-0699

学数据价值增值水平变化作为有效性检验的观测指标。通过 Vensim PLE 对构建的科学数据价值增值影响因素系统流图进行仿真。考虑到科学数据价值增值过程并不是一蹴而就的，而是需要在各因素的作用下持续推进，因此，本文设置仿真时间周期为 12 个月，设定时间步长为 1 个月，科学数据价值增值水平变化趋势如图 4 所示。科学数据价值增值水平在前期维持较低增长水平，从第 8 个月开始科学数据价值增值水平趋势由平缓上升变为快速上升趋势，即意味着从目前原始数据质量、数据存缴、数据组织与整合、数据共享和开发、数据政策与监管、资金保障 6 个维度入手进行科学数据管理优化，那么到第 8 个月时，科学数据价值增值水平将显现出优化效果。从当前国内外知名科学数据中心，如中国的国家科学数据中心、美国的开放科学数据中心等，实际运行情况来看，科学数据价值的增值主要体现在数据规模、质量、应用价值方面的提升，以科学数据中心为代表的主体，其数据价值增值过程涉及原始数据生产、数据存缴、数据保存、数据组织、数据整合、数据开放共享、数据开发等关键活动，这些活动的优化本身需要一定时间，继而产出相应的效果也需要一定周期，这与前期科学数据价值增值水平较低相呼应。随着以原始数据质量提高，及数据存缴、数据组织与整合、数据共享与开发等流程体系的不断成熟、科学数据政策与监管体系的不断完善，科学数据价值增值效能开始显著。科学数据价值的不断提升，进一步促进科学数据管理过程完善，实现科学数据价值增值的良性循环。从仿真结果可以看出，科学数据价值增值过程与实际情况相符，表明模型能够有效、真实地反映科学数据价值增值实现的动态变化过程。

5.4.2 情景分析

本研究通过改变模型中原始数据质量、数据存缴、数据组织与整合、数据共享和开发、数据政策与监管、资金保障等 6 个主要参数进行灵敏度分析。在不改变其他变量参数的情况下，将 6 个主要参数分别设置为 0.3、0.5、0.9，仿真得出各影响因素相同变化量时，科学数据价值增值水平的变化趋势，从而可以看出不同

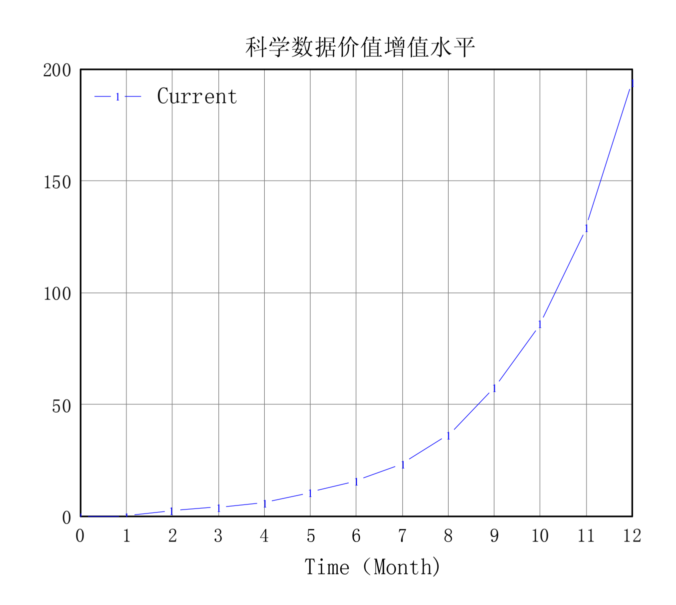


图 4　系统动力学模型有效性分析

Fig.4　Effectiveness analysis of the system dynamics model

影响因素对科学数据价值增值水平的影响程度。根据仿真结果（图 5），6 个影响因素对科学数据价值水平的影响均为增加量。其中原始数据质量对科学数据价值水平影响最大，其次是数据存缴、数据组织与整合、数据共享与开发等因素，而数据政策与监管、资金保障等辅助支撑因素同样对科学数据价值产生正向作用，但随着仿真时间推移，其影响力占比逐渐减小。

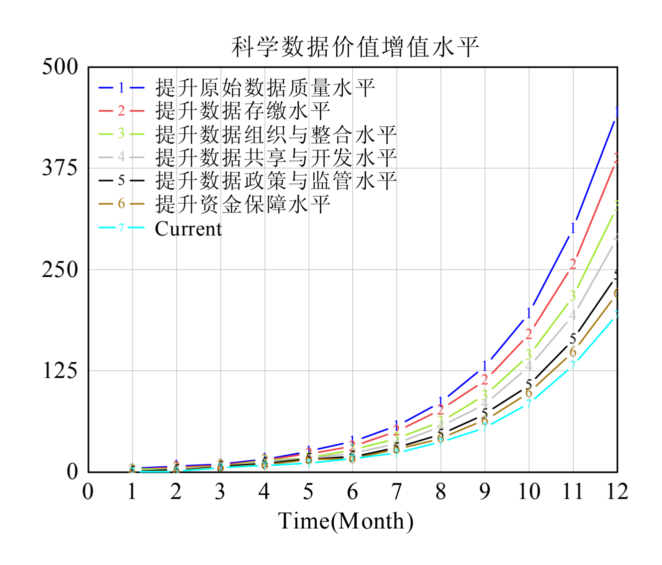


图 5　各影响因素对科学数据价值水平的变动效应

Fig.5 The effect of each influencing factor on the value level of scientific data

（1）科学数据收集因素。由图 6 可见，原始数据质量子系统中按影响程度由高到低依次为科研人员数据素养、数据软硬件设施、机构数据服务质量。科研人员数据素养水平对原始数据质量的影响占主导。其次，科学数据软硬件设施作为保障科学数据质量的物质基础，对科学数据质量影响显著。机构数据服务质量水平提升也可以显著提高原始数据质量。如图 7 所示，数据存缴子系统中，数据汇交意愿对科学数据存缴水平的影响最明显，其次是数据审核机制、数据存储与维护机制。但由于科研人员数据汇交意愿的提升，以及数据存储相关机制产生作用需要一个积累过程，所以前期数据存储水平变化不显著，从第 8 个月开始数据存储水平才不断提升，并且随着科研人员数据共享和汇交意愿的不断提高，数据存缴相关机制不断成熟完善，数据存缴水平增长速率不断提升。

（2）科学数据组织与开发因素。观察图 8 可以发现，在数据组织整合子系统中，元数据质量对数据组织与整合水平的正向促进作用更为显著，数据处理能力次之，数据融合能力促进作用相对最小。总体而言，随着仿真时间的推进，科学数据中心等科学数据共享机构的元数据质量，以及数据处理能力、数据融合能力的累积增强，科学数据组织与融合水平从第 7 个月起增长迅速，呈指数级增长。数据共享与开发子系统中，数据开放规模、数据共享平台质量、数据开发能力、数据共享服务能力对科学数据价值增值产生正向影响，其中影响程度最大的是数据开放规模，数据共享服务能力影响程度最小。而数据开放风险对科学数据价值增值产生反向抑制作用，数据开放风险越高，数据共享与开放水平越低。图 9 可见，仿真时间的前 8 个月，数据共享与开发子系统中各因素影响并不显著，

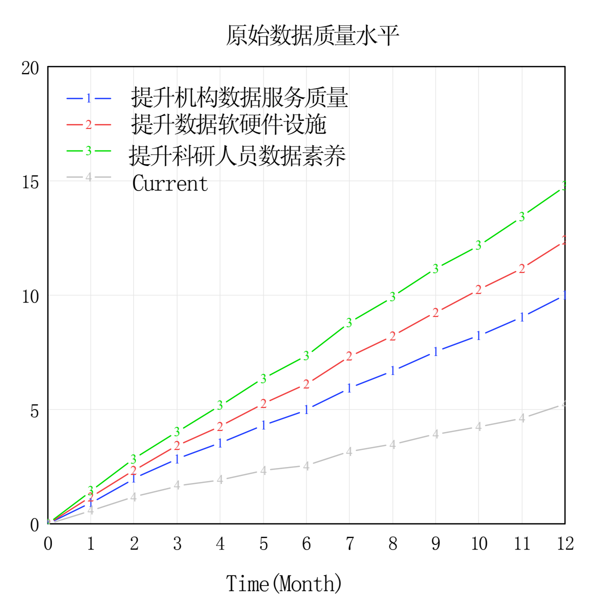


图 6　原始数据质量子系统关键因素影响趋势

Fig.6 Trends of key factors in the raw data quality subsystem

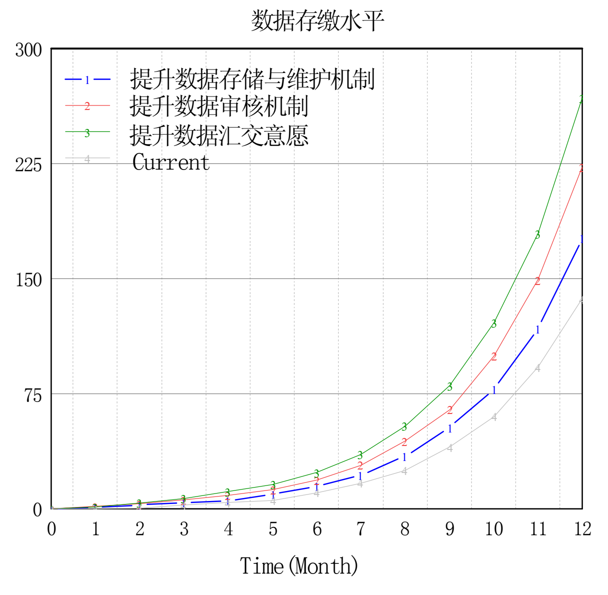


图 7　数据存缴子系统关键因素影响趋势

Fig.7 Trends of key factors in the data storage subsystem

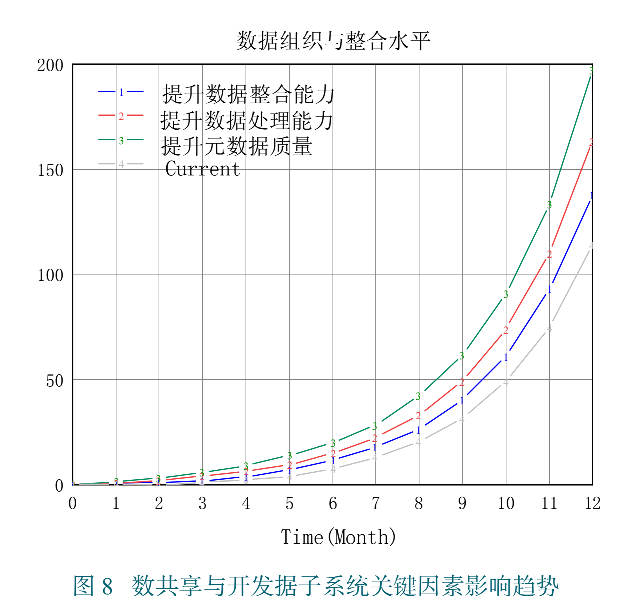


图 8　数共享与开发据子系统关键因素影响趋势

Fig.8 Influencing trend of key factors in data sharing and development subsystem

DOI：10.13998/j.cnki.issn1002-1248.23-0699

8 个月之后各因素产生的作用逐渐明显，这是因为这些因素对数据共享与开发水平的影响不是立竿见影的，需要一段时间积累之后才能显示出显著变化。

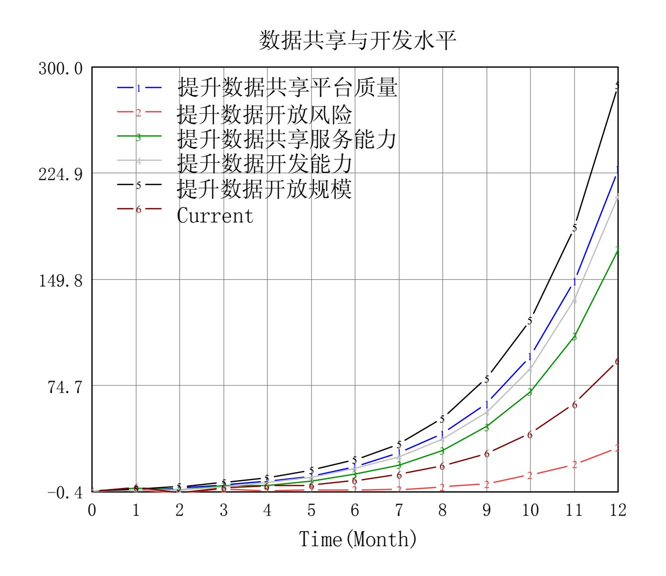


图 9　数据共享与开发子系统关键因素影响趋势

Fig.9　Trends of key factors in the data sharing and development subsystem

（3）保障因素。图 10 仿真结果可见，数据政策与监管子系统中，关键因素的影响程度从高到低依次为数据政策完善程度、数据评价与激励机制、数据监管机制。其中，数据政策完善程度对数据政策与监管水平影响最为显著。数据评价与激励机制、数据监管机制的影响程度区别不大。图 11 仿真结果可见，提升财政支持力度对资金保障水平的作用见效更快，数据价值增值收益的增长需要一段时间的发展积累方能显现，但长期来看，数据价值增值收益因素更具有可持续性。

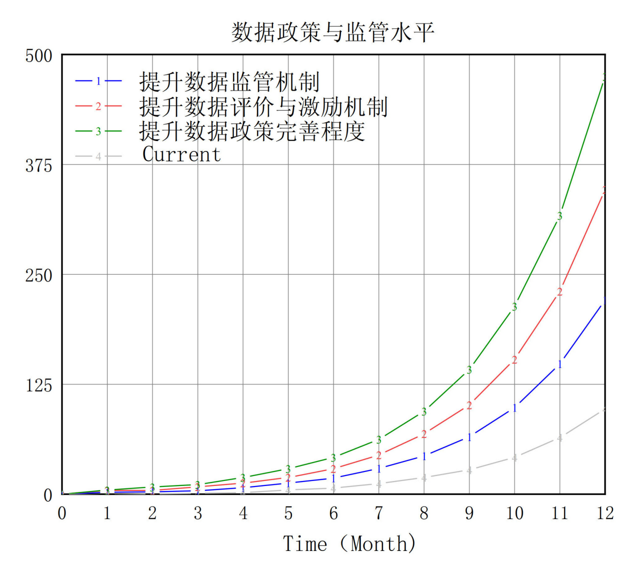


图 10　政策与监管子系统关键因素影响趋势

Fig.10　Influencing trend of key factors in the policy and regulatory subsystem

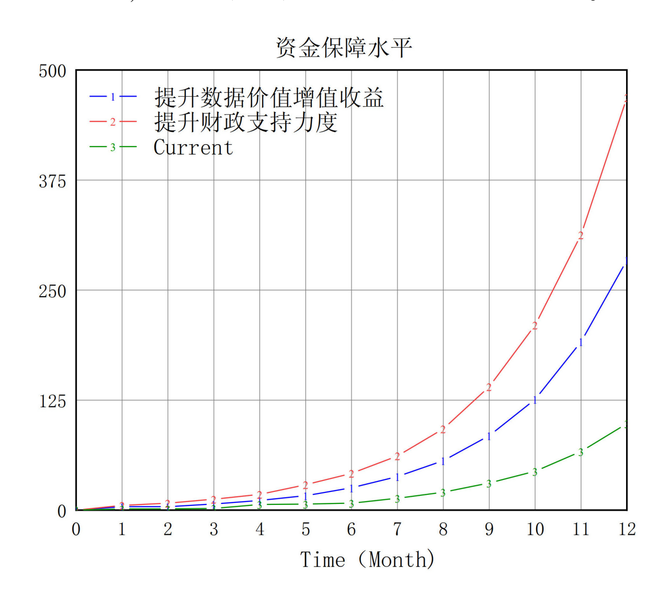


图 11　资金保障子系统关键因素影响趋势

Fig.11　The influencing trend of key factors in the fund guarantee subsystem

6　结论与讨论

本研究在文献收集与专家访谈获得科学数据增值过程的全面资料后，通过扎根理论这一成熟的质性研究方法逐级归纳得出科学数据价值增值影响因素理论模型，在此基础上进行系统动力学仿真，得出以下结论：①原始数据质量是科学数据价值增值的前提。科研人员为代表的数据生产者数据素养对科学数据质量影响最为显著，科研人员的数据收集和处理能力、数据分析能力、数据伦理意识等方面的提升，可以显著提高原始数据质量。②数据存缴水平对科学数据价值增值影响显著。当科研人员数据存缴的感知努力、感知风险降低，同时伴随科学数据共享政策的压力，将会促进科研人员提高存缴意愿，进而显著提升数据存缴规模效应。③科学数据组织与整合是科学数据价值增值形成的关键。总体上对科学数据组织与整合水平

影响最为显著的是元数据质量。元数据质量是科学数据加工、分类、融合的基础，元数据质量越高越有助于科学数据组织与整合水平的提升。④科学数据共享与开发是科学数据价值增值实现的关键，是科学数据价值增值实现的最后一步。在此过程中，科学数据开放规模、科学数据共享平台质量以及开发能力都对科学数据共享与开发水平产生积极促进作用。

本研究全局性地、动态地对科学数据增值过程和关键因素加以揭示。运用系统动力学模型从系统的角度出发，考虑了科学数据增值过程相关因素之间的相互作用和联系，避免了现有相关研究无法用实证数据衡量的局限性。同时，为开放科学数据治理实践提供了优化科学数据价值的明确目标。诚然，本研究虽通过专家调研、文献调研等方式收集了丰厚的质性数据，但科学数据相关实践随着软硬件环境迅速发展，具有动态性、复杂性，因此，今后需要进一步运用多元化方法持续跟踪收集科学数据价值增值的影响因素，对本研究结论进一步验证和完善。

7 启　示

基于仿真结果，本研究针对科学数据价值增值提出以下策略启示。

（1）培育数据素养，改善数据环境，提升科学数据质量。原始科学数据质量，是科学数据价值增值的基础。原始数据质量子系统中，科研人员的数据素养很大程度地影响了原始科学数据质量。因此，需要加强对科研人员的数据素养教育，提升科研人员获取、管理、利用数据的能力。高校、科研院所、图书馆等主体作为开展数据素养教育的主阵地，可以通过科学数据素养通识教育、学科专题数据素养教育等教育模式，采用慕课、在线课程、培训、咨询服务等丰富形式为科研人员提供便利高效的数据素养教育。科学数据基础设施是科学数据价值增值的重要支撑，科学数据的生成与组织、存储、利用都依赖完善的数据环境。应注重完善科学数据基础设施建设，立足科学数据用户需求，完善科学数据平台功能，丰富相关数据处理软件、数据存储服务，同时应遵循 FAIR 原则，聚焦基础设施的规范化、互操作以及可扩展性。

（2）着力动机激发，聚焦机制完善，增强科学数据存缴水平。在科学数据存储子系统中，科研人员以及科学数据管理机构的数据汇交意愿和科学数据审核机制的影响程度较大。因此，可以倡导与宣传提高科学数据共享氛围使科研人员提高主观规范；尽快建立科学数据共享的价值认同，以及科学数据汇交激励政策，提升科研人员数据共享的感知利益。从而以激发科研人员数据汇交动机为着力点，突破当前科学数据共享困境。同时，科学数据中心等开放数据管理和服务机构应完善审核机制，制定科学数据质量标准和相关业务规则，监管汇交至科学数据中心的数据质量，进而在科学数据存储环节提高科学数据价值。

（3）关注元数据质量，着力数据整合，优化科学数据组织和整合方式。元数据是科学数据组织和整合的关键因素，具备高质量元数据的科学数据更容易被组织和利用，因此，提升元数据质量是使科学数据价值增值的重要手段。当前科学数据元数据标准多局限于科学研究或科学管理的某一阶段，缺乏通用的科学数据元数据标准。需要面向科学数据开放共享、多领域互操作等利用需求，在核心元数据的基础上结合学科领域特点合理规划科学数据的元数据标准；元数据元素需提高完整性，应覆盖科学数据从产生到组织、整合、开发利用的全过程，使数据利用者了解科学数据生命周期过程。同时，需提供配套的元数据服务，例如，图书馆等科学数据管理服务机构可以提供在线编辑元数据的免费工具，便于科研人员便捷地生成高质量元数据。在对科学数据规范化标引与描述的基础上，建立多维度、多领域、多形态科学数据知识关联，促进开放科学环境下跨学科跨领域的科学创新。

（4）扩大开放规模，优化开放服务，提升科学数据开发利用能力。科学数据中心等科学数据管理和服务机构应扩大科学数据开放规模，包括数量规模、学科领域以及更丰富的元数据范围、开放接口等等，让更多科学数据可发现、可访问、可互操作是科学数据实现价值增值的前提。在此基础上，优化科学数据开

DOI：10.13998/j.cnki.issn1002-1248.23-0699

放服务，建设集开放、监管、服务、交易、交换多功能于一体的科学数据平台，积极探索和培育科学数据相关利益主体互惠共生模式，开展科学数据的深度挖掘和增值应用，促使科学数据价值最大化。可以通过需求导向或利益驱动等多种方式，利用云计算、人工智能等新兴数字技术，在保障国家数据安全和相关主体知识产权的前提下，探索科学数据产品和增值服务的市场化模式，顺应科学数据作为生产要素参与市场分配的新趋势。例如可以以江苏省等长三角区域为试点，充分整合区域科技、科研数据资源，建设一体化科学数据平台，配套激励政策吸引机构、数据科学家、社会公众积极参与数据产品化开发，鼓励应用创新；明确参与主体的权利与业务规范，促进区域科学数据实现商业化流通。

参考文献：

[1] HOLDREN J P. Increasing access to the results of federally funded scientific research[EB/OL]. [2023-07-06]. https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf.

[2] The European Parliament and the Council of the European Union. Directive (EU) 2019/1024 of the European parliament and of the council of 20 June 2019 on open data and the re-use of public sector information[EB/OL]. [2023-07-06]. https://www.legislation.gov.uk/eudr/2019/1024/introduction.

[3] Data FAIRport. Find, access, interoperate & re-use data[EB/OL]. [2023-12-06]. https://www.datafairport.org/.

[4] 温亮明, 李洋, 郭蕾. 国内外开放科学的实践进展与未来探索[J]. 图书情报工作, 2021, 65(24): 109-122.

WEN L M, LI Y, GUO L. The practice progress and future exploration of open science at home and abroad[J]. Library and information service, 2021, 65(24): 109-122.

[5] University of Cambridge. Data champions[EB/OL]. [2021-04-03]. https://www.data.cam.Ac.uk/intro-data-champions.

[6] 刘莉, 司莉. 科学数据治理实践：内容体系与发展趋势[J/OL]. 情报理论与实践, 2023: 1-9.(2023-07-24). https://kns.cnki.net/kcms/detail/11.1762.G3.20230724.1603.002.html.

LIU L, SI L. Scientific data governance practices: Content systems and development trend[J/OL]. Information studies: Theory & application, 2023: 1-9. (2023-07-24). https://kns.cnki.net/kcms/detail/11.176 2.G3.20230724.1603.002.html.

[7] 董诚, 侯敏. 科技资源共享价值最大化的三层次模型(VAA)[J]. 科技管理研究, 2013, 33(11): 231-234.

DONG C, HOU M. A three-layered model that sharing maximizes the value of scientific and technical resources (VAA)[J]. Science and technology management research, 2013, 33(11): 231-234.

[8] 李海舰, 赵丽. 数据价值理论研究[J]. 财贸经济, 2023, 44(6): 5-20.

LI H J, ZHAO L. Research of data valuation[J]. Finance & trade economics, 2023, 44(6): 5-20.

[9] ALBINO J P. An approach to value creation in open data for small and medium-sized enterprises using "R" ecosystem[C]// 2017 Computing Conference. Piscataway, New Jersey: IEEE, 2018: 346-351.

[10] KŘEMEN P, NEČASKÝ M. Improving discoverability of open government data with rich metadata descriptions using semantic government vocabulary[J]. Web semantics: Science, services and agents on the world wide web, 2019, 55(C): 1-20.

[11] 李旭晖, 凡美慧. 大数据中的知识关联[J]. 情报理论与实践, 2019, 42(2): 68-73, 107.

LI X H, FAN M H. Knowledge interaction in big data[J]. Information studies: Theory & application, 2019, 42(2): 68-73, 107.

[12] 马费成, 吴逸姝, 卢慧质. 数据要素价值实现路径研究[J]. 信息资源管理学报, 2023, 13(2): 4-11.

MA F C, WU Y S, LU H Z. Research on the path to realize the value of data elements[J]. Journal of information resources management, 2023, 13(2): 4-11.

[13] 阳巧英, 夏义堃. 我国数据要素价值形成机理、影响因素与实现路径——基于扎根理论的分析[J]. 图书与情报, 2023(2): 12-22.

YANG Q Y, XIA Y K. The formation mechanism, influencing factors and implementation path of data factor value[J]. Library & information, 2023(2): 12-22.

[14] 邓君, 贾晓青, 马晓君, 等. 科学数据价值鉴定标准研究[J]. 情报科学, 2013, 31(9): 37-41.

DENG J, JIA X Q, MA X J, et al. Study in appraisal exandard of scientific data value[J]. Information science, 2013, 31(9): 37-41.

[15] 顾立平, 陈新兰, 张潇月, 等. 开放科研数据中的数据价值提升策略[J]. 图书馆论坛, 2020, 40(9): 115-124.
GU L P, CHEN X L, ZHANG X Y, et al. A strategy for enhancing the value of open research data[J]. Library tribune, 2020, 40(9): 115-124.

[16] 孙建军, 李阳. 科学大数据: 范式重塑与价值实现[J]. 图书与情报, 2017(5): 20-26.
SUN J J, LI Y. Scientific big data: Paradigm remodeling and value realization[J]. Library & information, 2017(5): 20-26.

[17] 冯媛. 科学数据开放共享的价值共创模型及运行机制研究[J]. 图书馆, 2022(9): 29-37.
FENG Y. Value co-creation model and operation mechanisms of open sharing of scientific data[J]. Library, 2022(9): 29-37.

[18] 任颖, 李楠. 科学数据价值共创系统构建及仿真分析[J]. 数字图书馆论坛, 2023, 19(5): 42-53.
REN Y, LI N. Construction and simulation analysis of scientific data value co-creation system[J]. Digital library forum, 2023, 19(5): 42-53.

[19] 任福兵, 孙美玲. 基于价值链理论的政府开放数据价值增值过程与机理研究[J]. 情报资料工作, 2021, 42(4): 56-63.
REN F B, SUN M L. Research on the value-added process and mechanism of government open data based on value chain theory[J]. Information and documentation services, 2021, 42(4): 56-63.

[20] 宋姗姗, 白文琳. 中国大数据治理研究述评[J]. 农业图书情报学报, 2022, 34(4): 4-17.
SONG S S, BAI W L. A review of big data governance research in China[J]. Journal of library and information science in agriculture, 2022, 34(4): 4-17.

[21] GLASER B G. The grounded theory perspective: conceptualisation contrasted with description[M]. Mill Valley: Sociology Press, 2001: 145.

[22] MYERS M D, NEWMAN M. The qualitative interview in IS research: Examining the craft[J]. Information and organization, 2007, 17(1): 2-26.

[23] 郭安元. 基于扎根理论的心理契约违背的影响因素及其作用机制研究[D]. 武汉: 武汉大学, 2015.
GUO A Y. A study on the factors and mechanism of psychological contract violation based on the grounded theory[D]. Wuhan: Wuhan University, 2015.

[24] 柯江林, 孙健敏, 李永瑞. 心理资本: 本土量表的开发及中西比较[J]. 心理学报, 2009, 41(9): 875-888.
KE J L, SUN J M, LI Y R. Psychological capital: Chinese indigenous scale's development and its validity comparison with the western scale[J]. Acta psychologica sinica, 2009, 41(9): 875-888.

[25] 蒋畅和, 刘祖德, 赵云胜. 基于系统动力学的安全经验仿真试验研究[J]. 安全与环境工程, 2012, 19(4): 97-101.
JIANG C H, LIU Z D, ZHAO Y S. Simulation study on safety experience based on system dynamics[J]. Safety and environmental engineering, 2012, 19(4): 97-101.

[26] 李旭. 社会系统动力学: 政策研究的原理、方法和应用[M]. 上海: 复旦大学出版社, 2009.
LI X. Social system dynamics: Principles, methods and applications of policy research[M]. Shanghai: Fudan Press, 2009.

[27] 高晓宁, 胡威, 臧国全. 科研数据共享效率影响因素系统动力学仿真分析[J]. 情报理论与实践, 2022, 45(8): 146-153, 103.
GAO X N, HU W, ZANG G Q. System dynamics simulation analysis on factors affecting sharing efficiency of scientific research data[J]. Information studies: Theory & application, 2022, 45 (8): 146-153, 103.

[28] 袁红, 王焘. 政府开放数据生态系统可持续发展实现路径的系统动力学分析[J]. 图书情报工作, 2021, 65(17): 13-25.
YUAN H, WANG T. System dynamics analysis of the sustainable development path of the government open data ecosystem[J]. Library and information service, 2021, 65(17): 13-25.

DOI：10.13998/j.cnki.issn1002-1248.23-0699

Influencing Factors of Scientific Data Value Increment Based on System Dynamics

SUN Lili[1], WANG WeiJie[2], SHENG Jiefei[3]

(1. Institute of Information Management and Technology, Nanjing University of Technology, Nanjing 210009; 2. Communication University of China, Nanjing 211172; 3. School of Economics and Management, Nanjing University of Technology, Nanjing 211816)

Abstract: [Purpose/Significance] To explore the influencing factors of the added value of scientific data and reveal the inherent development laws of the added value of scientific data. [Method/Process] First, based on the interview data of 18 experts and the research results related to the value appreciation of scientific data in existing literature, the grounded theory method was adopted. Through open coding, main-axis coding, and selective coding, 19 categories, 6 main categories, and 3 core categories were finally obtained. From this, a theoretical model of the factors influencing the value appreciation of scientific data was obtained. On this basis, the Vensim PLE tool was used to establish a dynamic model of the value appreciation system of scientific data, and the process of value appreciation of scientific data was dynamically simulated and analyzed to reveal the relationship between various influencing factors and the value appreciation of scientific data. [Results/Conclusions] In the process of increasing the value of scientific data, the quality factor of raw data is a prerequisite, and high-quality raw scientific data are conducive to the integration and secondary development of subsequent scientific data. The data literacy of data producers represented by researchers has the most significant impact on the quality of scientific data. The level of data storage and payment has a significant impact on the added value of scientific data. When the perceived effort and perceived risk of data storage and payment by researchers decrease, accompanied by the pressure of scientific data sharing policies, researchers become more likely to increase their willingness to save and pay, thereby significantly improving the scale effect of data storage and payment. The organization and integration of scientific data is the key to the formation of value-added scientific data. Overall, the metadata quality has the most significant impact on the level of scientific data organization and integration. The quality of metadata is the foundation of scientific data processing, classification, and integration. The higher the quality of metadata, the more it helps to improve the level of scientific data organization and integration. The sharing and development of scientific data is the key to realizing the value-added of scientific data. It is the final step in realizing the value-added of scientific data. In this process, the scale of scientific data openness, the quality of scientific data sharing platforms, and development capabilities all have a positive promoting effect on the level of scientific data sharing and development.

Keywords: scientific data; value increment; influencing factors; system dynamics; data elements; FAIR